图书情报工作
第63卷 第5期 2019年3月

专利引用关系形成的解释框架:一个指数随机图模型视角*

杨冠灿[1] 陈亮[2] 张静[2] 李纲[3]
[1] 中国人民大学信息资源管理学院 北京 100872 [2] 中国科学技术信息研究所 北京 100038
[3] 武汉大学信息管理学院 武汉 430072

摘要:[目的/意义]近年来,围绕着专利引文网络结构特征的研究出现大量的研究成果,这些成果都从某种程度上折射出专利引文关系的形成受到来自属性特征之外关系特征的影响,而现有的以回归方法为基础的统计推断方法难以将这些因素纳入到分析框架中,因此,急需探索新的方法。[方法/过程]从关系形成视角,专利引用关系形成可表示三种广义的关系形成过程:自组织影响过程、自身属性影响过程、网络协变量影响过程,并建立关系形成过程与网络配置间的映射关系,最终,形成一整套可用于理解复杂专利引用关系形成问题的解释框架。[结果/结论]提出一整套可用于理解复杂专利引用关系形成问题的解释框架,该框架是未来进一步构建网络统计模型的理论基础,另外,解释框架包含丰富的网络配置项,预示着未来指数随机图模型在文献计量、科学网络分析上广阔的应用前景。

关键词:专利引用关系形成 解释框架 统计网络模型
分类号:G306
DOI:10.13266/j.issn.0252-3116.2019.05.012

1 引言

专利引文由于能够追踪技术发展的脉络,测量国家、区域间的技术扩散、技术溢出,衡量发明、技术的质量与价值,分析创新主体的技术战略行为[1],因此,在科技评价过程中具有十分重要的作用。近年来,学者们将网络分析方法引入专利引文分析,涌现了大量专利引文网络相关的研究成果,这些研究极大丰富了专利引文分析的视角,突破了传统单纯依赖专利引文频数进行分析的思路,采用可视化及描述性统计方法对专利引文的结构、动态特征开展了大量的讨论[2,3]。然而,由于缺乏一套综合性的统计网络推断方法,极大限制了研究人员从社会网络分析中获得有效的分析结论[4]。

专利引文网络的形成是一个复杂问题,其影响因素可能包括专利引文网络自身演化过程、专利自身属性特征以及网络外部因素等;传统的统计推断方法(如回归方法)是以属性型数据为基础的,以独立性假设为前提的。而网络分析的核心对象是关系数据,对其设定独立性假设是不合适的,这也使得传统的回归分析方法不具备对上述假设进行检验的能力。

因此,建立一套能够适应网络研究复杂性的统计网络推断方法成为当前专利引文网络研究中亟待解决的关键问题,而解决问题需要跳出传统以属性数据为基础的分析框架。指数随机图(Exponential Random Graph,ERG)模型是一种以关系形成(tie formation)为对象的研究方法[5],ERG模型是以关系数据为基础,以依赖性假设为条件,选择网络局部结构作为网络统计量来观察复杂网络的整体结构特征,从而获得对于网络复杂性、关联性以及随机性整体认知的方法[6],本文的研究目标是:在ERG模型的指引下,梳理现有文献中关于专利引文网络形成影响因素的关键研究进展,建立起关键研究进展与ERG模型中网络局部结构间的映射关系,形成一个可用于指导专利引用关系形成模型的解释框架,为后续网络统计推断打下基础。

本文依据如下顺序进行组织:第二部分是文献综

* 本文系国家自然科学基金项目“基于指数随机图模型的专利引用关系形成影响因素及机理研究”(项目编号:71403256)和国家自然科学基金项目“面向专利文本中实体关系抽取的远程监督方法研究”(项目编号:71704169)研究成果之一。
作者简介:杨冠灿(ORCID:0000-0002-1706-1884),讲师,博士,E-mail:yanggc@ruc.edu.cn;陈亮(ORCID:0000-0002-3235-9806),副研究员,博士;张静(ORCID:0000-0003-0291-0959),副研究员,博士;李纲(ORCID:0000-0001-5573-6400),教授,博士,博士生导师。
收稿日期:2018-07-23 修回日期:2018-11-15 本文起止页码:100-109 本文责任编辑:徐健

述，主要是对当前文献中论述的关于专利引用关系形成影响因素及机理进行梳理；第三部分提出一个面向专利引用关系形成的统计网络解释框架，该框架由网络自组织影响、专利自身属性影响、网络协变量影响三部分构成；最后，对解释框架进行总结，列举可能的应用场景，展望后续研究方向。

2 专利引用关系形成的影响因素

2.1 专利引用关系形成问题

专利引用关系的形成问题并不是一个新问题，专利引用关系形成与专利引文网络分析本质是一个问题，都是对于专利引文网络结构特征进行探索。但两者观察的视角有所区别，专利引文网络分析则主张网络的宏观结构决定了个体关系的形成，因此，采用整体网、自我中心网络等结构测量方法，从描述统计视角观察网络结构特征对专利引文网络的影响；而专利引用关系的形成是从微观视角出发来认知整体网络结构特征，它遵循社会选择过程(social selection process)理论的基本假定——社会网络结构特征是由个体关系累积形成的局部网络子结构所决定的，具体表现为：网络的形成是局部涌现的结果；网络关系形成不仅受网络内在自组织及属性的影响，还受到外部因素的影响；多个社会过程是同时发生的等[7]，以上述假定为基础，学者提出应更多考虑利用统计推断方法实现对社会网络结构特征深刻理解，ERG 模型就是一种主流的统计网络方法[8]。因为存在这种观察视角上的差异，使得对专利引用关系形成的研究更多是关注于网络结构的统计推断问题。由于专利引文网络分析发展更早，其研究中形成的理论、方法也成为当前整个网络视角下研究专利引文问题的基础，因此，在术语使用上，本文并不严格区分这两个概念。

2.2 专利引用关系形成影响因素及机理的探索

专利引用关系形成的影响因素是一个复杂问题，影响引用关系形成的因素既包括专利自身的属性，也包括专利引文网络的自组织过程，甚至有时还受到外部网络因素的影响。

专利引文分析的早期发展阶段主要关注发明者属性对专利引用关系形成的影响[9](尤其是高被引专利)：专利引文(patent citation)频次的研究[10－12]是引文研究的起点，通过相关实证研究确立了专利的属性因素——被引频次对于专利引用关系形成的影响；研究人员发现时间因素也在很大程度上影响专利引用关系的形成，尤其是专利被引频次的计算效果，如“时间截面”(time truncation)、“引证膨胀”(citation inflation)以及引文时滞(citation lag)问题[13]；研究人员通过观察，认为地域因素也会影响到专利引用关系的形成，如专利引文频次与地域、技术领域以及语言的关系研究[14]；专利发明人、申请人的属性特征对于专利引用关系的形成会产生影响，如专利自引与它引[15]、申请人来源类型(如大学或科研机构)[16－17]、H 指数[18]、专利引用非专利文献(Non-patent literature)[19]。

随着研究的深入，特别是专利引文网络分析方法的发展，学者们开始更多关注网络自身结构特征对专利引用关系形成的影响。研究人员从专利引文网络自身视角观察到专利引文网络存在多个维度：包括专利的直接引用网络[20－22]、耦合网络[23－24]、共引网络[25]、间接引用网络[26－27]；验证了一系列适用于专利引文网络的微观测量指标：如中心度、密度、集聚系数等；证实了专利引文网络也符合特定网络结构特征，如符合小世界特征[28]、幂律分布[29]、富人俱乐部(Rich-club)[30]、直径收缩(shrinking diameters)[31]、网络中介与闭合结构[32]等网络特征等。

随着专利引文网络的复杂性逐步成为学界的共识，学者们发现专利引用关系形成很大程度上受网络外部因素的影响，如受社会交流结构(social structure)和技术交流结构(intellectual structure)共同影响[33]。也有学者研究了专利引文网络和专利耦合、共引网络以及间接引用网络之间的相互影响关系[34－35]，引文网络反映的知识流动与地理距离之间存在一定关联作用[36－37]，并提出利用这种基于引用关系的多重关系协同网络机制，能够提高专利引文分析的精度，弥补遗失专利引文(missing links)[38]，绘制更完整的技术前沿[39－40]。

2.3 传统统计推断方法的失灵

一直以来，学者们尝试将统计推断方法应用到专利引文分析、专利网络分析过程中，希望利用统计推断结论来指导研究。早期的针对专利引文的统计推断方法是建立在属性数据以及独立性假设基础上。在方法上，主要选择回归分析，如 Logistic 回归[41]、Logit 回归[13,42]以及零膨胀泊松(Zero-inflated Poisson)[43]等方法。在理论假定上也对于专利引文研究做了诸多限定，1981 年 L. C. Smith 归纳引文的 5 项基本假设的第五项就是：“假定所有专利都是均等的”[44]；专利引文分析阶段最核心的专利指标就是专利被引频次指标，被引频次忽略了引用关系、属性、外部因素等其他的相关因素，其实质就是将专利频次视为每一个专利发明

人对专利质量独立的投票。正如 B. Yoon 在 2004 那篇经典文章中批评的那样:“专利分析仅显示专利引用对之间的直接关系,忽略了全部专利之间的整体关系”[45]。上述方法的问题在于:回归分析是独立性假设为前提的,要求多个自变量之间相对独立,不能高度线性相关。否则,就会存在“多重共线性”问题,使得变量的显著性检验失去意义,模型的预测功能失效。而专利引文网络分析的基本单元是专利之间的关系,因此,对于关系型数据设置独立性假设是不适合的。

随着网络分析方法的发展,出现了一些解决关系变量、属性与关系变量之间相关关系检验的方法,如二次指派程序(Quadratic Assignment Procedure,QAP),该方法研究一个网络与多个网络之间的相似性问题,从一定程度上解决关系数据之间的相关性评价[46],另外,该方法也可以通过一定形式的网络转化将属性数据包含进来,早期的网络统计分析模型如 P1 模型也是采用类似的思路。然而,这种方法无法应用更复杂的情景中,如基于加权和二值属性特征的混合测量。因此,缺乏能涵盖多方面影响因素的综合统计分析框架限制了专利引文网络分析方法的发展。

3 建立专利引用关系形成解释框架

3.1 指数随机图(ERG)模型

ERG 模型是一种以关系形成为对象的研究方法,其发轫于 1959 年 P. Erdos 和 A. Renyi 提出的社会网络统计分析模型(伯努利图分布),随后,W. Holland 和 L. Samuel 提出了二元关系模型,即 模型,该模型是关于关系发生概率与行动者的“扩张性”与“聚敛性”之间关系的统计模型,以指数形式进行表达;1986 年,O. Frank 和 S. David 引入了马尔科夫依赖,1996 年 S. Wasserman 于将上述模型扩展成为可以包含图中任何统计配置的 ERGM/ 模型,1999 年 J. Anderson 提出了对上述模型的参数化估计方法使得模型有了重要的进展[47],目前,ERG 模型正处于快速发展期。

ERG 模型是一个可以根据研究内容进行调整的扩展模型,其最一般的形式为:

$$\Pr(Y=y)=\left(\frac{1}{k}\right)exp\{\sum_{A}\eta_{A}g_{A}(y)\}$$

其中,求和是包含所有的配置 A 的加总,η_A 是对应的配置 A 的参数,$g_A(y)=\prod_{yij\in A}y_{ij}$是对应配置的网络统计量,是标准化常数,确保公式为适当的概率分布[48]。

ERG 模型是在对有序的局部网络结构进行观察基础上的建模,通过特定的参数估计过程,局部网络结构所对应的参数值可以被计算出来,从而实现对于复杂网络结构的统计推断过程。而在整个过程中,对于领域学者而言,关键就是要将一组具有理论意义的概念、假设、算法、指标等特征从现有文献中提取出来,建立与 ERG 模型中局部网络结构之间的映射关系。

3.2 专利引用关系基本特征

专利引文网络是由专利之间的引用关系所构成的网络,该网络区别与其他网络诸如合作网络的特征在于:通常而言,引用关系具有无权性、单向性、截面性三种基本特征。无权性(unweighted)特征是指:专利引用关系是将施引专利和被引专利联系起来的纽带,这种联系是二值化的,即仅存在引用与不引用两种关系,一般不考虑一项专利被另一项专利多次引用的情形;有向性(directed)特征主要是由专利引用行为存在时序所造成的,一项专利通常只会被在其公开时点后的专利所引用,且仅能引用在其公开时点前的专利。因此,专利引用关系是单向的,且该方向是逆时序的;截面性(cross-section)又可以称为截面数据,即引文网络是对某一时点前全部数据的截面展示。专利引用关系由于具有上述无权、单向性、截面的特征,使得整个专利引文网络具有了区别于发明人合作网络、专利文本相似网络的结构特征,具体表现在网络中二元组、三元组、以及高阶结构的具体构成,也影响到了对应的配置、网络统计量等因素,也进一步影响到了解释框架的构成。见图 1。

3.3 专利引用关系形成解释框架

以专利引用关系的基本特征为基础,结合现有文献研究的进展以及关系形成理论的基本思想,笔者对影响专利引用关系形成可能的关键效应进行组合,形成一个面向于统计网络推断的解释框架(见图 2),该框架由三部分组成:网络自组织影响过程、自身属性影响过程以及网络二元关系协变量影响过程。这三部分又分别对应具体的网络配置以及网络效应,这里网络配置(configurations)是社会网络结构中代表局部规律性的一个可能的局部子图,ERG 模型通过将上述模式化的局部子图转化为网络统计量的过程来实现对整个网络的统计推断,可见,网络配置是整个 ERG 建模过程关键的一个环节;而网络效应(effects)是希望通过 ERG 模型检验的网络配置与关系形成之间的因果关系。整个解释框架的核心就是要通过搭建现有专利引文网络研究的理论与 ERG 模型之间的桥梁,为后续的网络统计推断提供依据。

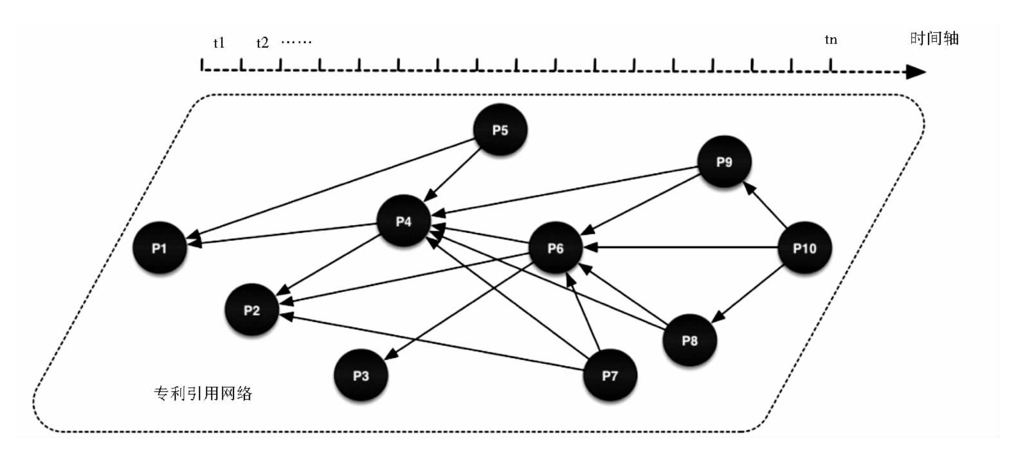


图 1　专利引用网络示意

图 2 显示，解释框架由三部分构成，包括网络自组织、属性因素以及网络二元关系协变量。网络自组织（Network Self-Organization）影响过程是指引用关系在某些情境下自发的可以促使其他引用关系的形成，通过累积逐渐形成一个有序的网络结构。这一过程是网络的内生效应，是纯粹产生于网络关系内在的运动变化过程；属性因素影响过程强调网络个体属性特征对于关系网络的形成具有影响，网络中各专利间在属地、领域、范围、质量、影响力上是存在差异的，这些差异与网络自组织特征一起影响着专利引用关系形成。除此之外，专利引文网络的外部的因素可以通过网络的二元关系协变量（dyadic covariates）的形式影响专利引用关系的形成[49]，如专利之间的语义关联、地理距离、人际关联因素也会从外部影响专利引用关系的形成，网络二元关系协变量虽然与网络数据有着相似的结构，但其关注的中心是网络本身的协变量效应。因此，一个完整的解释专利引用关系形成的框架由上述三部分构成[7]。

专利引用关系形成解释框架

	网络自组织过程					专利自身属性过程				网络协变量过程
效应及解释	稀疏效应	扩张效应	聚敛效应	连通效应	传递效应	发送效应	接受效应	同质效应	异质效应	协同效应
	关系数量或网络密度对于关系形成的影响	网络节点出度分布对于关系形成的影响	网络节点入度分布对于关系形成的影响	2-路径构造对于关系形成的影响	传递闭合结构对于关系形成的影响	链出节点属性对于关系形成的影响	链入节点属性对于关系形成的影响	具有相同属性的节点对于关系形成的影响	具有相异属性的节点对于关系形成的影响	网络二元协变量对于关系形成的影响
网络配置	弧	出-K-星	入-K-星	2-路径	传递闭合	发送者	接受者	趋同性	趋异性	伴生性
理论及假设	稀疏引文网络	倾向链接 引证膨胀 专利耦合	倾向链接 高被引 同被引	结构洞 网络中介 遗失链接 间接引文	网络闭合 聚集性	地域倾向 领域倾向 学科倾向 机构倾向 审查员类型倾向性	地域倾向 领域倾向 学科倾向 机构倾向 审查员类型倾向性	地域趋同 领域趋同 学科趋同 机构趋同 地理趋同	地域趋异 领域趋异 学科趋异 机构趋异 地理趋异	社会结构 知识结构 语义网络 地理距离

图 2　专利引用关系形成解释框架

需要说明的是，图 2 中的解释框架是以有向、无权、截面数据为假定条件的，因此，在针对具体复杂专利引用关系形成问题时还需要根据具体进行一定的调整。然而，笔者所提出的解释框架已经涵盖了当前影响专利引用关系形成的主要因素，例如引文稀疏性问题、引文倾向链接、结构洞、传递性、网络闭合、地域（机构、学科）同质性或异质性、知识网络（语义网络）与引文网络的相关性问题等。一方面，ERG 模型还处于快速发展期，各种软件包及工具（如 Pnet 和 Statnet）已经提供几百种的网络配置[50]；同时，对于专利引文网络的研究也在不断进展，会有新的理论算法会被纳入进来。实际上，ERG 模型的生命力就体现在其扩展性上。另外，在图 2 中，有些效应对应两种网络配置，例如聚敛效应，之所以会产生这样的情形，是由于真实网络是嵌套的，同一种效应在二元组、三元组、以及高阶结构的观察视角下，具体的网络配置形式会有所差别，在实际建模过程中，需要根据具体的研究目标来确定网络配置。

3.4　网络自组织影响过程

网络自组织过程是网络研究中的一个主要研究分支。ERG 模型下，与专利引用关系形成相关的网络自组织效应主要有以下 5 种：稀疏效应、扩张效应、聚敛效应、连通效应、传递效应。如图 3 所示：

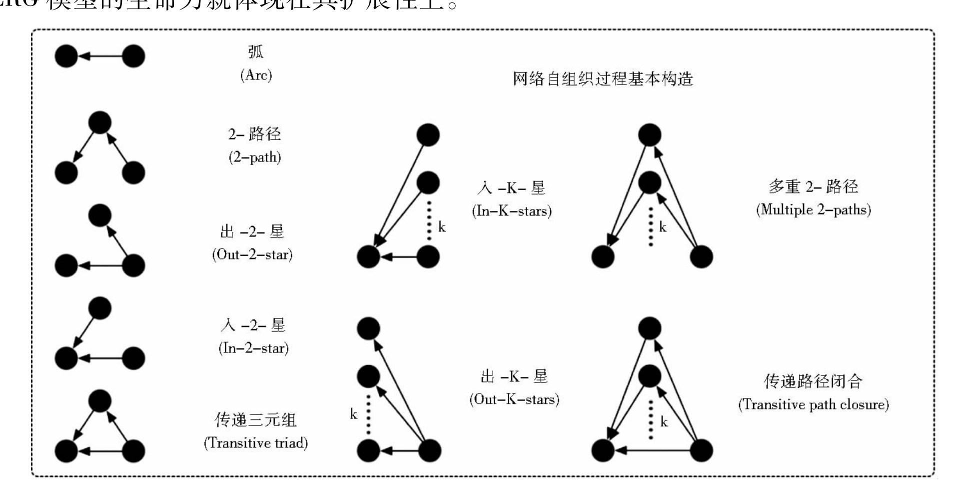


图 3　专利引文网络自组织效应所对应的网络配置

3.4.1　稀疏效应　稀疏效应是指假定网络中其他因素不变的条件下，引文关系数量对于专利引用关系形成的影响。就专利引文的数量而言，以往的文献主要归纳了两种特征：E. Yan 提到专利引文网络是极为稀疏的网络[35]；A. B. Jaffe 提出专利引文数量随着时间推移存在“引证膨胀”的现象[15]。虽然同时存在上述两点特征，但稀疏性往往会表现的更为突出，随时间变化，专利引文网络稀疏特征的增长规模（增长专利数的平方）要远大于专利引文增长的数量[31]。因此，通常网络中的关系数量效应对于专利引用关系形成的影响是较小的，甚至是负向的。之所以要考虑关系数量效应，主要是该效应是简单随机图产生的依据，代表了网络中最大不确定性的范围，是比较研究的基础。在 ERG 模型中，测量这种稀疏效应所对应的网络配置见图 3 弧（arc）。

3.4.2　扩张效应　扩张效应（popularity effect）是指假定网络中其他因素不变的条件下，节点出度分布对专利引用关系形成的影响。扩张效应所对应的网络配置是一个星形结构，从中心节点链接出两条或者多条弧，具体到专利引文网络中，表示某一专利参考引用其他多项专利的行为，该行为所对应的扩张效应可以用节点出度来测量；同时，从三元组视角观察，扩张效应所对应的网络配置类似专利耦合关系结构[2]，而从枢纽与权威视角观察，扩张效应所对应的网络配置也可以用来解释枢纽特征[30]。因此，上述两种专利引文网络中的现象也可以利用扩张效应所对应网络配置进行测量。图 3 中，ERG 模型中测量这种扩张效应的网络配置被称为“出-2-星”（Out-2-star）或者是“出-K-星”（Out-K-stars）。

3.4.3　聚敛效应　聚敛效应（activity effect）是指假定网络中其他因素不变的条件下，节点入度分布对于专利引用关系形成的影响。聚敛效应所对应的网络配置是一个星形结构，从中心节点链接入两条或者多条弧，具体到专利引文网络中，少量专利会被其他专利高频引用，而有些专利则很少被引用，这种特征在社会网络中通常被称为“富人俱乐部”、“马太效应”或者“倾向

链接”现象,而上述问题正是聚敛效应要测量的内容。同时,从三元组视角观察,聚敛效应所对应的网络配置类似专利同被引结构[24],也类似于枢纽与权威研究中的权威特征[30]。因此,上述两种专利引文网络中的现象也可以利用聚敛效应所对应网络配置进行测量。具体到ERG模型中,测量这种聚敛效应的网络配置被称为“入-2-星”(In-2-star)或者“入-K-星”(In-K-stars)。

3.4.4 连通效应 连通效应是指假定网络中其他因素不变的条件下,2-路径配置对于专利引用关系形成的影响。连通效应所对应的2-路径配置(无论是简单的还是多重2-路径配置)是一种较为特殊的局部配置,一方面处于中心的节点接受从其他节点链入弧,同时它也链向其他节点。具体到专利引文网络中,从关系视角来看,它体现了两篇专利文献之间经由中间专利文献的联系形成了一种间接连通的关系,它类似于专利间接引文结构,也可以用于表示潜在“遗失链接”[38],同时,从节点视角来看,处于两篇专利文献中心位置的专利文献又扮演着中介者的作用[32]。以上述两点研究为基础,2-路径配置还可以作为专利引文网络“结构洞”特征测量的手段。见图3,ERG模型中测量这种连通效应的网络配置被称为“2-路径”(2-path)或者“多重2-路径”(Multiple 2-path)。

3.4.5 传递效应 传递效应是指假定网络中其他因素不变的条件下,传递闭合对于专利引用关系形成的影响。传递闭合(transitivity closure)也是专利引文网络研究中一种较受关注的结构,但在传统专利引文网络文献中更多的表现为对聚集系数或者网络聚集特征的测量。然而,更全面地考虑传递闭合结构,其特点还表现为两个方面:①传递闭合是在2-路径配置基础增加的一条弧,该弧的增加使得“遗失链接”显性化,配置内部的关系更为稳健,该特征可以用于分析专利技术的演化路径[32];②在传递闭合配置中,度分布并不均匀,某些节点具有更多的入度,而这种在传递闭合配置中的入度优势要优于单纯聚敛效应配置中的入度优势。因此,传递效应也可用于识别知识流动过程中的源头[51]。见图3,ERG模型中测量这种传递效应的网络配置被称为“传递三元组”(transitive triad)或者“多重传递闭合”(multiple transitive closure)。

3.5 专利自身属性影响过程

专利自身的属性特征对于专利引用关系的形成也是具有十分重要的影响。相关研究显示专利在地域、专利申请人、专利发明人、所涉技术广度与深度、审查员类型、技术领域以及对于科学知识的依赖程度方面都存在较大的差异,这些与专利自身属性相关的因素也会在不同程度上影响了专利引用关系的形成。

3.5.1 发送者效应(或接受者效应) 在专利引文网络中,某些属性特征能够通过影响网络中其他专利参与引用行为的程度对专利引用关系的形成产生影响。对于有向网络而言,这种效应可分为两种:一种是接受者效应(receiver effect),指假定网络中其他因素不变的条件下,专利自身所具备的某种属性能够影响该专利被其他专利更多的引用。例如,在美国申请的专利由于其在公开程度、专利质量以及受市场关注程度上有别于他国专利。因此,如果一项专利具备“在US申请”的属性,那么,该专利就可能会被更多的引用[52]。该配置如图4中的接受者所示。

另一种情况是发送者效应(sender effect),具体到专利中是指假定网络中其他因素不变的条件下,专利自身所具备的某种属性能够影响该专利会更积极地引用其他专利。相关研究认为美国专利局对专利引用行为规定了“举证责任”义务,如果专利申请人在申请过程中没有公开现有技术,在后期诉讼中将处于不利的地位,这样势必导致在美国专利局申请的专利的平均参考文献数量要多于其他国家[53]。对应在网络统计测量中,如果一项专利具备“在US申请”的属性,那么,该专利就可能倾向引用更多的其他专利。该配置如图4中的发送者所示。

3.5.2 同质效应(或异质效应) 除了考虑单方面的专利属性特征外,引用行为所涉及的双方专利的属性特征也会影响到专利引用关系的形成。这种影响因素可以被称之为同质效应(homogeneous effects)或者异质效应(heterogeneous effects)。同质效应是指假定网络中其他因素不变的条件下,引用行为所涉及的双方专利具有同样属性将会对专利引用关系的形成产生影响,该效应所对应的配置见图4中的趋同性所示;异质效应则是指引用行为所涉及的双方专利具有不同的属性会对专利引用关系的形成产生影响,该效应所对应的配置见图4中的趋异性所示。例如,有学者证明在美国专利数据中,在美国专利局申请的专利会优先引用来自美国专利局授权的专利[13],同样,学者也发现在欧洲专利局申请的专利也存在类似的现象[54]。根据上述相关文献,似乎说明存在基于共同专利受理国特征的专利引用倾向性;同样,在实际过程中,也可能会存在基于不同专利受理国特征的专利引用倾向性,这种倾向性就可以通过异质效应来检验。非专利引文问题也是专利引文研究中非常值得关注的问题,通常,

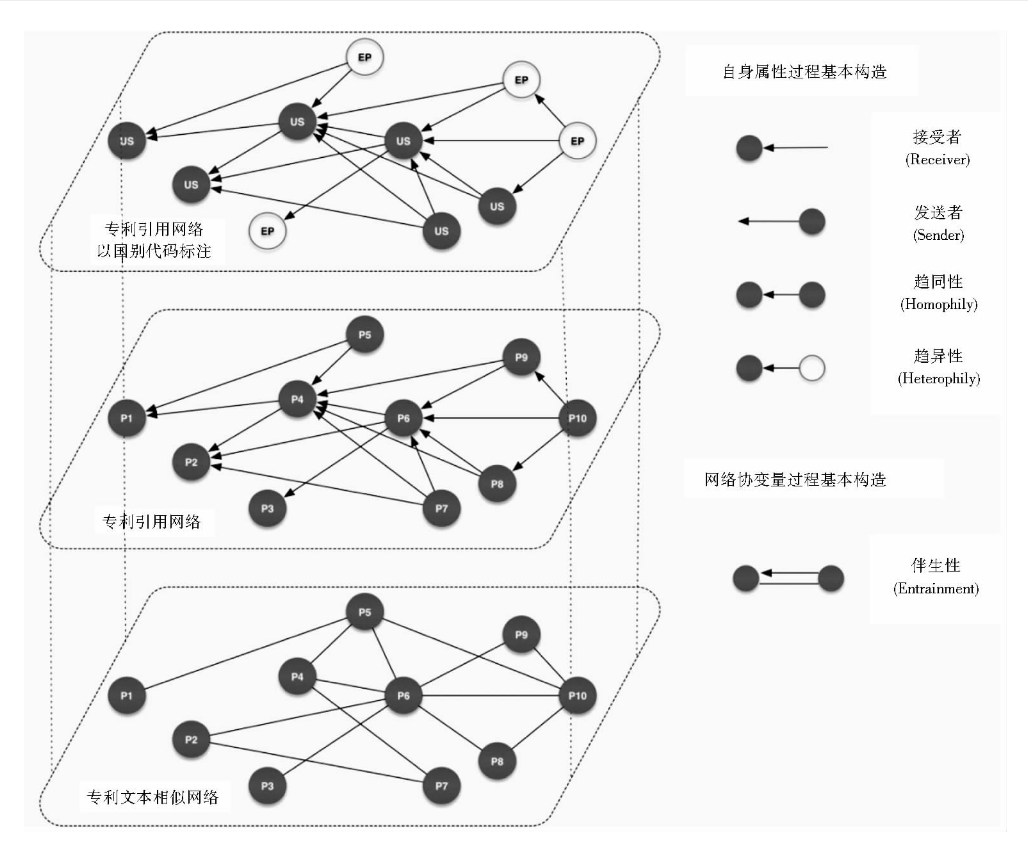


图 4 专利属性及网络协变量影响因素的配置示例

专利既可能引用专利文献也可能引用非专利问题，该问题的解决可以通过将专利审查员引文问题视为不同类型节点属性（专利文献、非专利文献）之间的关系问题，于是，就可以采用同质性效应或异质性效应来解决上述问题。

3.6 二元关系协变量影响过程

相关的研究表明，在专利引用行为过程中，专利权人的合作关系、专利发明人流动关系、专利间文本的语义相似性、专利间技术分类相似性以及地理距离等引用关系以外的特征也对专利引用关系的形成具有影响。

协同效应是指假定网络中其他因素不变的条件下，二元关系协变量网络表现出的与专利引用关系的伴生性对于专利引用关系形成的影响。该效应所对应的网络配置见图 4 中的伴生效应，举例而言，图 4 中专利引文网络中专利 P9 与专利 P10 之间存在引用关系，同时，在专利文本相似网络中，专利 P9 与专利 P10 之间存在本文相似关系，此时，这两种类型网络的关系就构成伴生关系。在具体网络分析中，伴生性的表现形式有很多种类，包括专利权人的合作关系、专利发明人流动关系、专利间文本的语义相似性、专利间技术分类相似性以及地理距离等，这些都影响到专利引用关系的形成。另外，“专利审查员引文也可以通过协同效应进行检验。专利引用关系可分为审查员引文、申请人引文等，于是，该问题可以被视为一个多类型关系问题（也可被称为多重关系、多层关系等），通过构建网络二元关系协变量，就可以检验审查员引文网络对于专利引用关系形成的影响，同时，由于审查员引文和申请人引文之间并不具有重叠性，因此，测量二元关系协变量的结果实际上就可以解释：在考虑网络中其他条件不变的前提下，相对于专利申请人引文，专利审查员引文的增多对专利引用关系形成影响的效果。”

4 应用与展望

从理论上而言，引用关系形成的解释框架是希望概括一个能够包含解释所有专利引文特征、子结构的体系，进而在该解释框架的指引下，指数随机图模型可以分别检验各种效应是否真实存在以及对于引用关系形成的效果。因此，理想上，解释框架应该能够包含所有影响专利引用关系形成的子结构、模式（如前所述是通过网络配置的构建实现）。当然在实际应用过程中，专利引文数据非常复杂，多重关系、异质信息、有权、时序等因素使得希望构建一个能够包罗万象的解释框架是不现实的。但即便如此，笔者提出的解释框架对于

专利引文关系——这一种特殊的网络系统还是具有较为全面的解释力。

如前所述,建立一个综合的统计推断模型对于专利引文网络具有重要的作用。以下将列举3个实例说明ERG模型如何能弥补现有专利引文网络研究的不足,利用统计网络推断为专利引文网络研究提供指导。

(1)针对复杂网络效应的推断。在专利引文网络中,高被引问题一直是研究的热点,但高被引问题是一个由多种复杂效应共同起作用的结果,如倾向链接、接受者效应、同质效应往往是同时存在的,如何判断复杂网络效应的效果是一个学术界关注的核心问题。倾向链接认为新的专利引用关系形成是基于现有专利引用关系的存在,尤其是入度分布特征;而接受者效应则认为专利接受更多的专利引用是由专利自身个体的属性因素决定的,例如专利权人所属的国籍、专利所涉新兴技术领域等原因,而同质效应则认为专利对的引用行为更倾向于在具有同类属性的专利对之间发生,例如专利权人同属于一个国籍等。一个相似实例是C. Zhang利用ERG检验了同质性、传递性、以及倾向链接对于科学合作关系形成的作用[55]。

(2)对于高度嵌套变量的推断。如文献综述部分所述,研究人员发现专利引文网络存在多个维度:包括专利的直接引用网络、耦合网络、共引网络、间接引用网络等,这些网络各自都在一定程度上对于专利引用关系形成产生了影响,但真实网络往往同时包含这些关系,于是,对于这种具备高度嵌套的关系如何进行检验尚没有好的方法。指数随机图模型的出现恰恰能够检验具备高度嵌套特征的网络变量的效应问题,事实上,G. Robins等提供了一套同时检验传传递性、连通性、度分布特征(聚敛或扩张效应)的测量方法[56]。

(3)多重关系、异质网络整合的权重设置问题。通过整合多重关系、异质网络实现对网络的多视角观察是近年来学术上的研究热点。在对多重关系、异质网络进行整合的过程中,权重设置问题成为一个难点。ERGM模型能够通过对模型进行参数估计获得各种因素对于整体网络形成影响的参数估计值,该估计值可以作为科学设置权重的解决思路,相关研究可以参考D. Lusher等在著作中第10章——“指数随机图模型的扩展:多重关系网络与二部网络”[7]。

目前,ERG模型的研究目前还处于起步阶段,随着研究的深入,针对更为复杂情况,如多重网络、动态网络的网络配置也正在被纳入到ERG模型中来。本文针对专利引用关系形成提出的解释框架也仅是一个探索性的框架,在未来,笔者将尝试将更多的现有理论成果引入ERG模型,通过统计网络推断过程获得对专利引用关系形成更为深刻的洞见。

参考文献:

[1] OECD. OECD Patent Statistics Manual[M]. Paris: OECD Publishing, 2009:14-16.

[2] YANG G C, LI G, LI C Y. Using the comprehensive patent citation network (CPC) to evaluate patent value[J]. Scientometrics, 2015, 105(3): 1319-1346.

[3] VAN RANN A F J. Patent citations analysis and its value in research evaluation: a review and a new approach to map technology-relevant research[J]. Journal of data and information science, 2017, 2(1): 545-38.

[4] SHUMATE M, PALAZZOLO E T. Exponential random graph (p*) models as a method for social network analysis in communication research[J]. Communication methods and measures, 2010, 4(4): 341-371.

[5] ROBINS G, PATTISON P, KALISH Y. An introduction to exponential random graph (p*) models for social networks[J]. Social Networks, 2007, 29(2): 173-191.

[6] GOODREAU S M, HANDCOCK M S, BUTTS C T. Statnet: software tools for the representation, visualization, analysis and simulation of network data[J]. Journal of statistical software, 2008, 24(1): 1-11.

[7] LUSHER D, KOSKINEN J, ROBINS G. Exponential random graph models for social networks[M]. New York: Cambridge University Press, 2012:16-28.

[8] HARRIS J K. An introduction to exponential random graph modeling[M]. London: SAGE Publications, 2013:1-13.

[9] JAFFE A B, DE RASSENFOSSE G. Patent citation data in social science research: overview and best practices[J]. Journal of the Association for Information Science and Technology, 2017, 68(6): 1360-1374.

[10] NARIN F, CARPENTER M P, WOOLF P. Technological performance assessments based on patents and patent citations[J]. IEEE Transactions on engineering management, 1984, EM-31(4): 172-183.

[11] TRAJTENBERG M. A penny for your quotes: patent citations and the value of innovations[J]. The RAND journal of economics, 1990, 21(1): 172-187.

[12] HALL B H, JAFFE A B, TRAJTENBERG M. The NBER patent citation data file: lessons, insights and methodological tools[R]. Cambridge: National Bureau of Economic Research, 2001.

[13] ALCÁCER J, GITTELMAN M, SAMPAT B. Applicant and examiner citations in U.S. patents: an overview and analysis[J]. Research policy, 2009, 38(2): 415-427.

[14] ALCÁCER J, GITTELMAN M. Patent citations as a measure of knowledge flows: the influence of examiner citations[J]. Review of

economics and statistics, 2006, 88(4): 774 – 779.

[15] JAFFE A B, TRAJTENBERG M. Patents, citations, and innovations: a window on the knowledge economy [M]. New York: MIT Press, 2002: 4 – 19.

[16] CRISCUOLO P, NARULA R, VERSPAGEN B. Role of home and host country innovation systems in R&D internationalisation: a patent citation analysis[J]. Economics of innovation and new technology, 2005, 14(5): 417 – 433.

[17] OWEN-SMITH J, POWELL W W. The expanding role of university patenting in the life sciences: Assessing the importance of experience and connectivity[J]. Research policy, 2003, 32(9): 1695 – 1711.

[18] GUAN J C, GAO X. Exploring the h - index at patent level[J]. Journal of the Association for Information Science and Technology, 2009, 60(1): 35 – 40.

[19] BRUSONI S, CRISCUOLO P, GEUNA A. The knowledge bases of the world's largest pharmaceutical groups: What do patent citations to non-patent literature reveal? [J]. Economics of innovation and new technology, 2005, 14(5): 395 – 415.

[20] CHANG S B, LAI K K, CHANG S M. Exploring technology diffusion and classification of business methods: using the patent citation network [J]. Technological forecasting and social change, 2009, 76(1): 107 – 117.

[21] GRESS B. Properties of the USPTO patent citation network: 1963 – 2002[J]. World patent information, 2010, 32(1): 3 – 21.

[22] STERNITZKE C, BARTKOWSKI A, SCHRAMM R. Visualizing patent statistics by means of social network analysis tools [J]. World patent information, 2008, 30(2): 115 – 131.

[23] HUANG M H, CHIANG L Y, CHEN D Z. Constructing a patent citation map using bibliographic coupling: a study of Taiwan's high-tech companies[J]. Scientometrics, 2003, 58(3): 489 – 506.

[24] EGGHE L, ROUSSEAU R. Co-citation, bibliographic coupling and a characterization of lattice citation networks[J]. Scientometrics, 2002, 55(3): 349 – 361.

[25] BOYACK K W, KLAVANS R. Co - citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? [J]. Journal of the Association for Information Science and Technology, 2010, 61(12): 2389 – 2404.

[26] HU X, ROUSSEAU R, CHEN J. On the definition of forward and backward citation generations[J]. Journal of informatics, 2011, 5(1): 27 – 36.

[27] ATALLAH G, RODRÍGUEZ G. Indirect patent citations[J]. Scientometrics, 2013, 67(3): 437 – 465.

[28] HUNG S W, WANG A P. Examining the small world phenomenon in the patent citation network: a case study of the radio frequency identification (RFID) network[J]. Scientometrics, 2009, 82(1): 121 – 134.

[29] BRANTLE T F, FALLAH M H. Complex innovation networks, patent citations and power laws[C]// 2007 Portland International Conference on Management of Engineering & Technology. Piscataway: IEEE, 2007: 540 – 549.

[30] SMILKOV D, KOCAREV L. Rich-club and page-club coefficients for directed graphs[J]. Physica A: statistical mechanics and its applications, 2010, 389(11): 2290 – 2299.

[31] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graphs over time: densification laws, shrinking diameters and possible explanations[C]// Proceeding of the eleventh ACM SIGKDD international conference. New York: ACM, 2005: 177 – 187.

[32] WANG J C, CHIANG C H, LIN S W. Network structure of innovation: can brokerage or closure predict patent quality? [J]. Scientometrics, 2010, 84(3): 735 – 748.

[33] WHITE H D, WELLMAN B, NAZER N. Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group[J]. Journal of the Association for Information Science and Technology, 2004, 55(2): 111 – 126.

[34] SHIBATA N, KAJIKAWA Y, TAKEDA Y. Comparative study on methods of detecting research fronts using different types of citation [J]. Journal of the Association for Information Science and Technology, 2009, 60(3): 571 – 580.

[35] YAN E, DING Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other[J]. Journal of the Association for Information Science and Technology, 2012, 63(7): 1313 – 1326.

[36] THOMPSON P. Patent citations and the geography of knowledge spillovers: evidence from inventor - and examiner-added citations [J]. Review of economics and statistics, 2006, 88(2): 383 – 388.

[37] BOTTAZZI L, PERI G. Innovation and spillovers in regions: evidence from European patent data[J]. European Economic review, 2003, 47(4): 687 – 710.

[38] CHEN D Z, HUANG M H, HSIEH H C. Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology[J]. Journal of informetrics, 2011, 5(3): 400 – 412.

[39] MORRIS S A, VAN DER VEER MARTENS B. Mapping research specialties[J]. Annual review of information science and technology, 2008, 42(1): 213 – 295.

[40] BRAAM R R, MOED H F, VAN RAAN A F J. Mapping of science by combined co - citation and word analysis. I. Structural aspects[J]. Journal of the Association for Information Science and Technology, 1991, 42(4): 233 – 251.

[41] CRISCUOLO P, VERSPAGEN B. Does it matter where patent citations come from? Inventor vs. examiner citations in European patents[J]. Research policy, 2008, 37(10): 1892 – 1908.

[42] BAKKER J. The log-linear relation between patent citations and

patent value［J］. Scientometrics, 2016, 110(2): 1－14.

［43］LEE Y G, LEE J D, SONG Y I. An in-depth empirical analysis of patent citation counts using zero-inflated count data model: the case of KIST［J］. Scientometrics, 2007, 70(1): 27－39.

［44］NICOLAISEN J. Citation analysis［J］. Annual review of information science and technology, 2007, 41(1): 609－641.

［45］YOON B, PARK Y. A text mining based patent network: analytical tool for high-technology trend［J］. The journal of high technology management research, 2004, 15(1): 37－50.

［46］LEE W J, LEE W K, SOHN S Y. Patent network analysis and quadratic assignment procedures to identify the convergence of robot technologies［J］. PloS One, 2016, 11(10): e0165091.

［47］SNIJDERS T A B, PATTISON P E, ROBINS G L. New specifications for exponential random graph models［J］. Sociological methodology, 2006, 36(1): 99－153.

［48］ROSE KIM J Y, HOWARD M, COX PAHNKE E. Understanding network formation in strategy research: exponential random graph models［J］. Strategic management journal, 2016, 37(1): 22－44.

［49］ROBINS G. Doing social network research［M］. London: SAGE, 2015:81－85.

［50］MORRIS M, HUNTER D R, HANDCOCK M S. Specification of exponential-family random graph models: terms and computational aspects［J］. Journal of statistical software. 2008, 24(4): 1548－7660.

［51］BATAGELJ V. Efficient Algorithms for citation network analysis［EB/OL］.［2017－12－06］. https://www.researchgate.net/publication/1956732_Efficient_Algorithms_for_Citation_Network_Analysis.

［52］DERNIS H, HOISL K, WEBB C. Analysing European and international patent citations［EB/OL］.［2017－12－06］. http://www.oecd-ilibrary.org/science-and-technology/analysing-european-and-international-patent-citations_883002633010.

［53］MEYER M. What is special about patent citations? Differences between scientific and patent citations［J］. Scientometrics, 2000, 49(1): 93－123.

［54］COCKBURN I M, KORTUM S, STERN S. Are all patent examiners equal? The impact of examiner characteristics［R］. Cambridge: National Bureau of Economic Research, 2002.

［55］ZHANG C, BU Y, DING Y. Understanding scientific collaboration: homophily, transitivity, and preferential attachment［J］. Journal of the Association for Information Science and Technology, 2018, 69(1): 72－86.

［56］ROBINS G, PATTISON P, WANG P. Closure, connectivity and degree distributions: exponential random graph (p*) models for directed social networks［J］. Social Networks, 2009, 31(2): 105－117.

作者贡献说明：

杨冠灿：负责论文思路框架构建、主体内容撰写、实验与结果分析；

陈亮：负责文献调研，参与实验数据采集、整理与实验结果分析；

张静：参与实验数据采集与结果分析；

李纲：负责论文思路、框架上提供指导和建议。

Framework for Explanations of Patent Citation Formation: An Exponential Random Graph Model Perspective

Yang Guancan[1] Cheng Liang[2] Zhang Jing[2] Li Gang[3]

[1] School of Information Resource Management of Renmin University of China, Beijing 100872

[2] Institute of Science and Technical Information of China, Beijing 100038

[3] School of Information Management of Wuhan University, Hubei 430072

Abstract: ［**Purpose/significance**］Although there have been efforts of scholars to answer the question, what's determinants of patent citation formation are not solved satisfactorily. Scholars find formation of patent citation is influenced by the structure characteristics of patent citation network. However, the current framework of statistical inference methods based on logistical regression is failing to incorporate the above factors, so an innovative method need to be introduced. ［**Method/process**］From a tie formation perspective, patent citation formation represents three broad category of tie formation processes: attribute-based processes, self-organizing network processes and covariates processes. Furthermore, based on these processes, the paper establishes a mapping relationship between those processes with particular types of configurations. Finally, a framework is proposed for understanding the complexity of patent citation formation. ［**Result/conclusion**］The paper introduces a framework for understanding patent citation formation, which lays the groundwork for statistical network modeling in the future. In addition, broadly network configuration selection from the framework offers significant opportunities to extend existing bibliometrics and open new pathways in complexity of scientific network analysis.

Keywords: patent citations formation framework for explanations statistical network analysis